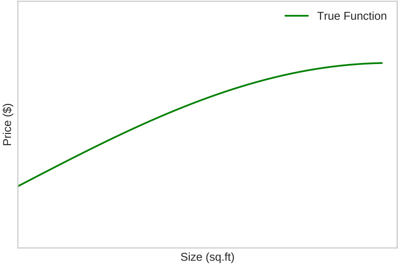
Bias-Variance

Nipun Batra and teaching staff

IIT Gandhinagar

July 29, 2025

Here, the true function $f_{\theta_{\text{true}}}$ is used to model the relation $y_t = f_{\theta_{\text{true}}}(x_t)$



Modeling the relation

This behavior varies due to training set randomness.

This behavior varies due to training set randomness.
Therefore, it is important to measure performance **averaged over all possible training sets** (of size N).

This behavior varies due to training set randomness.
Therefore, it is important to measure performance **averaged over all possible training sets** (of size N).

$$E_{\text{training set}}[\text{error of } \hat{\theta}(\text{training set})]$$

gives a measure of the average error by doing an expectation of the errors of all possible training sets of size N.

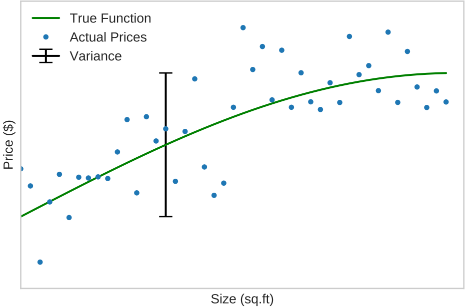
Therefore, $E_{train}[\text{at a point } x_t] = f(\text{noise, bias, variance})$

Formally defining the 3 sources of error: Noise

Noise is an **irreducible error** captured by the error term ϵ.
The equation of the relation becomes $y_t = f_{\theta(true)}(x_t) + \epsilon_t$

Formally defining the 3 sources of error: Noise

Noise is an **irreducible error** captured by the error term ϵ.
The equation of the relation becomes $y_t = f_{\theta(true)}(x_t) + \epsilon_t$
The noise is mean-centered around 0 with spread called the variance of the noise, denoted by σ^2.



Variance in the noise

Formally defining the 3 sources of error: Noise

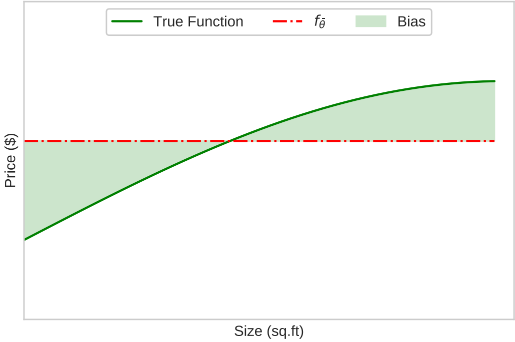
Noise is an **irreducible error** captured by the error term ϵ.
The equation of the relation becomes $y_t = f_{\theta(true)}(x_t) + \epsilon_t$
The noise is mean-centered around 0 with spread called the variance of the noise, denoted by σ^2.
That is, it can be denoted by $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$

Formally defining the 3 sources of error: Bias

Bias is a measure of how flexible the fit is in capturing the true function $f_{\boldsymbol{\theta}_{\text{true}}}(x)$

$$\text{Bias}(x_t) = f_{\boldsymbol{\theta}_{\text{true}}}(x_t) - f_{\bar{\boldsymbol{\theta}}}(x_t)$$

where $f_{\bar{\boldsymbol{\theta}}}$ denotes the average fit over all datasets.



Bias of the fit

Formally defining the 3 sources of error: Bias

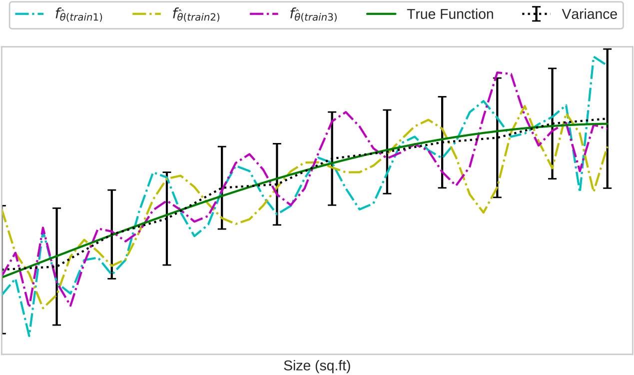
Bias is a measure of how flexible the fit is in capturing the true function $f_{\boldsymbol{\theta}_{\text{true}}}(x)$

$$\text{Bias}(x_t) = f_{\boldsymbol{\theta}_{\text{true}}}(x_t) - f_{\bar{\boldsymbol{\theta}}}(x_t)$$

where $f_{\bar{\boldsymbol{\theta}}}$ denotes the average fit over all datasets.
As $f_{\bar{\boldsymbol{\theta}}}$ denotes the average fit over all datasets, it can be expressed by $f_{\bar{\boldsymbol{\theta}}}(x_t) = E_{\text{train}}[f_{\hat{\boldsymbol{\theta}}}(x_t)]$

Formally defining the 3 sources of error: Variance

Variance of the fit is a measure of the variation in the fits when trained across different training sets.



Variance of a fit

Formally defining the 3 sources of error: Variance

Variance of the fit is a measure of the variation in the fits when trained across different training sets.
Variance of the fit can be defined by

$$\mathrm{var}(f_{\hat{\boldsymbol{\theta}}}(x_t)) = E_{\mathrm{train}}[(f_{\hat{\boldsymbol{\theta}}}(x) - f_{\bar{\boldsymbol{\theta}}}(x_t))^2]$$

where $f_{\hat{\boldsymbol{\theta}}}(x) - f_{\bar{\boldsymbol{\theta}}}(x_t)$ denotes the deviation that a specific fit has from the average.

Deriving Expected Prediction Error

Now we will see how,
$E_{train}[\text{at a point } x_t] = \sigma^2 + [\text{bias}(f_{\hat{\theta}}(x_t))]^2 + \text{var}(f_{\hat{\theta}}(x_t))$
where,
given a training set, the parameters $\hat{\theta}$ of the fit are learned as $f_{\hat{\theta}}$
and, the prediction at a point x_t for the model trained on that training set is $f_{\hat{\theta}}(x_t)$

Deriving Expected Prediction Error

Prediction Error at a point x_t can be calculated using the squared loss function.

Prediction error at $x_t = (y_t - f_{\hat{\theta}(train)}(x_t))^2$

To find the "Expected Prediction Error" at a point x_t we average out the prediction error at that point over all possible learned models. This can be done by finding the expectation of prediction error for that point over all possible training datasets ($train$) and labels for that point (y_t).

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[((y_t - f_{\theta(true)}(x_t)) + (f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)))^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

$$= E_{train,y_t}[(a+b)^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

$$= E_{train,y_t}[(a + b)^2]$$

$$= E_{train,y_t}[a^2 + 2ab + b^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

$$= E_{train,y_t}[(a+b)^2]$$

$$= E_{train,y_t}[a^2 + 2ab + b^2]$$

(Using Linearity of Expectation)
$= E_{train,y_t}[a^2] + 2E_{train,y_t}[ab] + E_{train,y_t}[b^2]$......................(Eqn. 1)

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

(Since there is no dependence on training set)

$$= E_{y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

$$(\because \text{ there is no dependence on training set})$$

$$= E_{y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))^2}_{\epsilon_t^2}]$$

$$= E_{y_t}[\epsilon_t^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

$$(\because \text{ there is no dependence on training set})$$

$$= E_{y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))^2}_{\epsilon_t^2}]$$

$$= E_{y_t}[\epsilon_t^2]$$

$$= \sigma^2 (\text{By definition})$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

$$(\because \text{ there is no dependence on training set})$$

$$= E_{y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))^2}_{\epsilon_t^2}]$$

$$= E_{y_t}[\epsilon_t^2]$$

$$= \sigma^2 (\text{By definition})$$

$$E_{train,y_t}[a^2] = \sigma^2 \text{.................}(\text{Eqn. } 2)$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$(\because \epsilon_t \text{ and } (f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)) \text{ are independent})$$

$$= E_{train,y_t}[\epsilon_t] \times E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

($\because$ ϵ_t and $(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))$ are independent)

$$= \underbrace{E_{train,y_t}[\epsilon_t]}_{=0} \times E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

(By definition ϵ_t has mean 0)

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$(\because \epsilon_t$ and $(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))$ are independent)

$$= \underbrace{E_{train,y_t}[\epsilon_t]}_{=0} \times E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

(By definition ϵ_t has mean 0)

$$E_{train,y_t}[ab] = 0 ....(\text{Eqn. } 3)$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)$ is independent of y_t)

$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)$ is independent of $y_t)$

$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= MSE(f_{\hat{\theta}(train)}(x_t))$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)$ is independent of y_t)

$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= MSE(f_{\hat{\theta}(train)}(x_t))$$

$E_{train,y_t}[b^2] = MSE(f_{\hat{\theta}(train)}(x_t))$ .... (Eqn. 4)

Deriving Expected Prediction Error

From Eqn. 1, 2, 3 and 4, we get,

Expected prediction error at $x_t = \sigma^2 + MSE(f_{\hat{\theta}(train)}(x_t))$

Now, we will further simplify the MSE term into bias and variance.

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[((f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t)) + (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)))^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

$$= E_{train}[(\alpha + \beta)^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

$$= E_{train}[(\alpha + \beta)^2]$$

$$= E_{train}[\alpha^2 + 2\alpha\beta + \beta^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

$$= E_{train}[(\alpha + \beta)^2]$$

$$= E_{train}[\alpha^2 + 2\alpha\beta + \beta^2]$$

(Using Linearity of Expectation)
$= E_{train}[\alpha^2] + 2E_{train}[\alpha\beta] + E_{train}[\beta^2]$ ....(Eqn. 5)

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2]$$

Deriving Expected Prediction Error

$$
\begin{aligned}
E_{train}[\alpha^2] &= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2] \\
&= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2] \\
&= E_{train}[bias(f_{\hat{\theta}}(x_t))^2] \qquad \text{(By definition of bias)}
\end{aligned}
$$

Deriving Expected Prediction Error

$$
\begin{aligned}
E_{train}[\alpha^2] &= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2] \\
&= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2] \\
&= E_{train}[bias(f_{\hat{\theta}}(x_t))^2] \qquad \text{(By definition of bias)} \\
&= bias(f_{\hat{\theta}}(x_t))^2 \\
&(\because \text{ bias is not a function of training data})
\end{aligned}
$$

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2]$$

$$= E_{train}[bias(f_{\hat{\theta}}(x_t))^2] \qquad \text{(By definition of bias)}$$

$$= bias(f_{\hat{\theta}}(x_t))^2$$

($\because$ bias is not a function of training data)

$$E_{train}[\alpha^2] = bias(f_{\hat{\theta}}(x_t))^2 \text{ ....(Eqn. 6)}$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$
\begin{aligned}
& E_{train}[\alpha\beta] \\
& = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))] \\
& = E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]
\end{aligned}
$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias$_t$ is not a function of training data)

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias_t is not a function of training data)

$$= bias \times \left(E_{train}[f_{\bar{\theta}}(x_t)] - E_{train}[f_{\hat{\theta}(train)}(x_t)]\right)$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias_t is not a function of training data)

$$= bias \times \left(E_{train}[f_{\bar{\theta}}(x_t)] - E_{train}[f_{\hat{\theta}(train)}(x_t)]\right)$$

$$= bias \times (f_{\bar{\theta}}(x_t) - f_{\bar{\theta}}(x_t))$$

($\because$ $f_{\bar{\theta}}(x_t) = E_{train}[f_{\hat{\theta}(train)}(x_t)$)

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias_t is not a function of training data)

$$= bias \times \left(E_{train}[f_{\bar{\theta}}(x_t)] - E_{train}[f_{\hat{\theta}(train)}(x_t)]\right)$$

$$= bias \times (f_{\bar{\theta}}(x_t) - f_{\bar{\theta}}(x_t))$$

$E_{train}[\alpha\beta] = 0$..........................(Eqn. 7)

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - E_{train}[(f_{\hat{\theta}(train)}(x_t)])^2]$$

$$(\because f_{\bar{\theta}}(x_t) = E_{train}[(f_{\hat{\theta}(train)}(x_t)]\ )$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - E_{train}[(f_{\hat{\theta}(train)}(x_t)])^2]$$

$$(\because f_{\bar{\theta}}(x_t) = E_{train}[(f_{\hat{\theta}(train)}(x_t)]\ )$$

$$= variance(f_{\hat{\theta}}(x_t))$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - E_{train}[(f_{\hat{\theta}(train)}(x_t)])^2]$$

$$(\because f_{\bar{\theta}}(x_t) = E_{train}[(f_{\hat{\theta}(train)}(x_t)]\ )$$

$$= variance(f_{\hat{\theta}}(x_t))$$

$$E_{train}[\beta^2] = variance(f_{\hat{\theta}}(x_t)) \text{...............(Eqn. 8)}$$

Deriving Expected Prediction Error

From Eqn. 1 - 8, we get,

Expected prediction error at x_t

$= \sigma^2 + MSE(f_{\hat{\theta}(train)}(x_t))$

$= \sigma^2 + bias(f_{\hat{\theta}}(x_t))^2 + variance(f_{\hat{\theta}}(x_t))$